Are Information or Data Patterns Correlated with Consciousness?

David Gamez

Department of Informatics, University of Sussex, Brighton, UK
david@davidgamez.eu / +44 7790 803 368

This is the author-generated version of an article published in *TOPOI*.

The original publication is available at www.springerlink.com.

Abstract

Scientific research on consciousness is attempting to gather data about the relationship between consciousness and the physical world. The basic procedure is to measure consciousness through first-person reports, measure the physical world and look for correlations between these sets of measurements. While this work has focused on neural correlates of consciousness, it has also been proposed that information states in the brain might be linked to consciousness. This paper uses Floridi's distinction between dedomena, data and information to state this claim more precisely and suggests that the best starting point for this work is the potential correlation between data patterns and consciousness. Floridi's method of levels of abstraction is used to understand how data can be measured at different levels of abstraction in the brain and the paper examines a number of problems with this work, which could make it difficult to prove that data patterns are correlated with consciousness.

Keywords. Consciousness; correlates of consciousness; information; data; information integration; levels of abstraction.

1. Introduction

> ... to the extent that a mechanism is capable of generating integrated information, no matter whether it is organic or not, whether it is built of neurons or of silicon chips, and independent of its ability to report, it will have consciousness.
>
> (Tononi 2008, p. 237)

Scientific research on the correlates of consciousness is attempting to identify the relationship between consciousness and the physical world without making a premature commitment to any particular theory of consciousness. A number of potential correlates have been identified, including neural synchronization, recurrent connections, quantum states and electromagnetic waves.[1] It has also been proposed that the pattern of integration and differentiation in the brain's information states, known as 'information integration', could be linked with consciousness (Tononi 2004, 2008). Some algorithms have been developed to measure information integration and the theory has been tested in some preliminary experimental work.

This paper looks at the philosophical and experimental issues that need to be addressed by any information- or data-based approach to the correlates of consciousness. The first section provides background information about the scientific study of consciousness and Tononi's information integration theory of consciousness. Section 3 uses Floridi's (2009) work to distinguish information from dedomena and data. Section 4 argues that we should focus on potential correlations between data patterns and consciousness and explains how data can be measured at different levels of abstraction in the brain. Section 5 outlines the formidable problems that are faced by any attempt to scientifically prove that data patterns are correlated with consciousness independently of the substrate in which the data is held.

[1] Some of these potential correlates are described in a review paper by Tononi and Koch (2008).

2. The Scientific Study of Consciousness

2.1 Experiments on the Correlates of Consciousness

Since consciousness developed its modern meaning in the 17th Century (Wilkes 1988) it has been an important philosophical research topic, and elaborate thought experiments have been devised about the relationship between consciousness and the physical world. Although a variety of theories have been put forward and a great deal of effort has been expended, it can be argued that our understanding of consciousness has failed to advance much beyond Descartes. The situation has changed in recent years with the emergence of the scientific study of consciousness. This aims to gather data about correlations between consciousness and the physical world while suspending judgement about the metaphysical and philosophical debates. Scientific data about the correlates of consciousness has many practical applications - for example, it could help us to identify consciousness in brain damaged patients who cannot communicate, and inform us about the level of consciousness in infants and animals.[2]

In an experiment on the correlates of consciousness, the general procedure is to measure consciousness, measure the physical world and look for a relationship between these sets of measurements. The measurement of consciousness relies on the working assumption that consciousness is a real phenomenon that can be reported verbally ("I see a red hat", "I taste stale milk") or through other behaviour, such as pushing a button, pulling a lever, short term memory (Koch 2004) or the Glasgow Coma scale (Teasdale and Jennett 1974). The assumption that consciousness can be measured through behaviour enables us to obtain data about it without a precise definition. However, this reliance on external behaviour does limit consciousness experiments to systems that are commonly agreed to be conscious, such as a human or a human-like animal. It is not possible to carry out this type of experiment on non-biological systems, such as computers, because a computer's reports are not generally regarded as a measurement of consciousness (Gamez 2012a).

The physical world is measured to identify spatiotemporal structures that might be correlated with the measurements of consciousness. In this paper I will focus on the correlates of consciousness in the human brain, which will be defined in a similar way to Chalmers' (2000) definition of the total correlates of consciousness: [3]

> **D1**. The correlates of a conscious experience, e_1, is a minimal set of one or more spatiotemporal structures in the brain that are present when e_1 is present and absent as a collection when e_1 is absent. This will be referred to as a *CC set*.

The notion of a minimal set is intended to exclude features of the brain that typically occur at the same time as consciousness, but whose removal would not lead to the alteration or loss of consciousness. For example, the correlates of consciousness might have prerequisites and consequences (Aru et al. 2012; de Graaf et al. 2012) that typically co-occur with consciousness, but the brain would be conscious in exactly the same way if the minimal set of correlates could be induced without these prerequisites and consequences. Correlates defined according to D1 would continue to be associated with consciousness if they were extracted from the brain or implemented in an artificial system. 'Spatiotemporal structures' is a deliberately vague term that captures anything that might be correlated with consciousness, such as activity in brain areas, neural synchronization, electromagnetic waves, quantum events, etc.

---

[2] Some people have expressed scepticism about the idea that there could be correlates of consciousness or that it is possible to identify them in the brain (Molyneux 2010; Noë and Thompson 2004). Some of the methodological issues with the scientific study of consciousness are discussed by Gamez in 'The Measurement of Consciousness' (currently under review).

[3] Chalmers (2000) distinguishes the total neural basis from the core neural basis: "A total NCC builds in everything and thus automatically suffices for the corresponding conscious states. A core NCC, on the other hand, contains only the 'core' processes that correlate with consciousness. The rest of the total NCC will be relegated to some sort of background conditions required for the correct functioning of the core." (Chalmers 2000, p. 26). Block (2007) makes a similar distinction.

A	B	C	D	Consciousness
0	0	0	0	0
0	0	0	1	0
0	0	1	0	1
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	1
0	1	1	1	1
1	0	0	0	0
1	0	0	1	0
1	0	1	0	1
1	0	1	1	1
1	1	0	0	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	1

Table 1. Illustrative example of correlations that could exist between consciousness and spatiotemporal structures in a physical system. A, B, C and D are spatiotemporal structures in the brain, such as dopamine, neural synchronization or 40Hz electromagnetic waves. In this example these are assumed to be the only possible features of the brain. '1' indicates that a feature is present; '0' indicates that it is absent. D has no correlation with consciousness and {A,B} and {C} are CC sets.

CC sets can be identified by systematic experiments that consider all possible combinations of spatiotemporal structures in the brain (see Table 1). A contrastive methodology is typically used that compares the state of the brain when it is conscious and unconscious, or compares conscious and unconscious states within a single conscious brain - for example, using binocular rivalry (Logothetis 1998) or through the subliminal presentation of stimuli (Dehaene et al. 2001). Correlations can be tested by making predictions about the level and contents of consciousness, which can be compared with first-person reports.

While scientific work on consciousness has typically focused on neural correlates, there are several different types of spatiotemporal structure in the brain that could form CC sets. The most frequently discussed ones are as follows:

- *Physical*. A physical correlate of consciousness is a pattern in a specific physical structure, such as neurons or electromagnetic waves. The CC set consists of the pattern and the substrate in which the pattern is held. For example, synchronization patterns in biological neurons could be a physical CC set.
- *Computational*. Some cognitive scientists and philosophers have claimed that the mind is in a deep sense functional or computational. If this is correct, some of the brain's computations could form CC sets, and the type of computer that was used to execute the computations (desktop computer, Babbage Analytical Engine, biological brain, etc.) would be irrelevant.[4]
- *Informational*. In this article I am focusing on the possibility that consciousness could be linked to patterns of information in the brain. In this case the CC set would only consist of a spatiotemporal information pattern, and not the substrate in which the information pattern happened to be instantiated. This view was pioneered by Tononi (2004, 2008), whose information integration theory of consciousness is described in the next section.

2.2 The Information Integration Theory of Consciousness

According to Tononi (2004, 2008) consciousness is identical with a system's ability to integrate information. An *integrated* state is one that results from interactions between the system's components, and not from the components acting independently. In common with Shannon (1948), Tononi defines *information* as the reduction of uncertainty. Each state reduces uncertainty by ruling out all of the other possible states that the system could be in. The greater the number of other possible states, the greater the reduction of uncertainty and the more information one gains by knowing that the system is in a particular state. For example, each state in a million pixel digital camera

[4] While there has been a substantial amount of speculation about the connection between computations in the brain and consciousness, few attempts have been made to specify what it means for a brain to implement a computation. This is necessary if one wants to carry out experiments on the computational correlates of consciousness. Chalmers' (2011) combinatorial state automata account is arguably the best theory of implementation that has been put forward. However, it has a number of problems, which I have discussed elsewhere in "Are there Functional Correlates of Consciousness?", currently under review. A draft copy is available on request.

contains a large amount of information because the camera has $2^{1,000,000}$ possible states. But there is no integration because the photovoltaic cells are acting independently. On the other hand, a set of Christmas tree lights with a single switch has a high level of integration because each light reflects the state of the switch, but each state contains little information because the system only has two states. *Information integration* is the extent to which the states of a system exhibit both information *and* integration. The conscious brain has high levels of information integration because it can enter a very large number of states and each state results from the causal interactions between the neurons. Tononi claims that the brain's high level of information integration is identical with its high level of consciousness.

A number of algorithms for measuring information integration have been put forward. These include stateless Φ (Tononi and Sporns 2003), state-based Φ (Balduzzi and Tononi 2008), a time series measure (Barrett and Seth 2011), liveliness (Gamez and Aleksander 2011) and an algorithm based on graph theory (Hadley et al. 2012).[5] Closely related measures include neural complexity (Tononi et al. 1994) and causal density (Seth et al. 2006). Tononi's stateless Φ and state-based Φ algorithms identify an area of the system that is predicted to be linked to consciousness (the main complex)[6] and output a number, Φ, that is supposed to correspond to the level or amount of consciousness. Balduzzi and Tononi (2009) have also proposed how a high dimensional structure can be calculated that is hypothesised to correspond to the contents of consciousness. A summary of Balduzzi and Tononi's (2008) algorithm is given in the appendix to this paper.

Some preliminary experiments have been carried out to test the information integration theory of consciousness. Lee et al. (2009) applied Tononi and Sporn's (2003) algorithm to EEG data and showed that there was a higher level of information integration in the gamma band when subjects were conscious. In another set of experiments the brains of conscious and unconscious subjects were monitored with EEG while they were stimulated with TMS (Ferrarelli et al. 2010; Massimini et al. 2009). The EEG signals resulting from a TMS pulse were longer range and more differentiated when subjects were conscious, which matched the predictions of information integration theory. In more recent work Casali et al. (2013) developed a 'perturbational complexity index' (PCI) that uses a compression algorithm to measure the differentiation and integration of the EEG patterns that result from a TMS pulse. Casali et al. (2013) describe a number of experiments that demonstrate a reliable correlation between PCI and subjects' level of consciousness.

2.3 Information Patterns Could be Correlates of Consciousness

Much of the discussion about the information integration theory of consciousness has focused on the algorithms that are used to measure information integration (see previous section). Although I will use the output of Balduzzi and Tononi's (2008) algorithm as a key example of an information pattern that might be correlated with consciousness, the main focus of this paper is on the broader claim that *some* pattern of information in the brain could be a correlate of consciousness. Furthermore, while Tononi claims that information integration is identical with consciousness, this article will only consider the weaker claim that information patterns in the brain could be *correlated* with consciousness. These claims are summed up in the following hypothesis, which uses the terminology introduced in the previous section:

> **H1**. One or more CC sets consist solely of information patterns. The substrate in which the information patterns are held is not part of these CC sets. An information pattern that was shown to be correlated with consciousness in one substrate (for example, the brain), would continue to be correlated with consciousness if it was implemented in a different substrate, such as a silicon chip.

H1 is a radical claim that is a significant departure from current scientific practice. Science usually attempts to measure the behaviour of specific aspects of the physical world and formulates laws that only apply to these aspects of the physical world. It is not just the pattern that counts, but the presence of this pattern in part of the physical world. For example, Newton's theory of gravity describes how masses behave on a particular spatiotemporal scale. These equations have no power to produce effects on their own and they would produce incorrect results if they were applied to electrical charges. While research on the physical correlates of consciousness fits within current scientific practice, in H1 the information patterns are treated as objects of study in and of themselves, independently

---

[5] The terms stateless Φ and state-based Φ were introduced in a previous paper to distinguish Tononi's two measures of information integration (Gamez and Aleksander 2011).

[6] While Tononi and Sporns' (2003) algorithm identifies a single main complex, Balduzzi and Tononi's (2008) algorithm can potentially identify more than one main complex in a system. In this paper I will speak as if information integration algorithms only identify a single main complex.

of the substrate in which they happen to be instantiated. Each information pattern is considered to be more than just a regularity or description of part of physical world: it is something that has an objective existence of its own; something *in* the physical world that could be correlated with consciousness.

H1 can only be considered to be a scientific hypothesis if it can be experimentally tested (Popper 2002). The testing of H1 will require changes to the standard scientific method, and the rest of this paper will consider what modifications are necessary and how likely they are to succeed. If H1 could be shown to be true, this would have significant implications for the types of system that we believe to be capable of consciousness. For example, an information pattern that was correlated with consciousness in the brain would continue to be correlated with consciousness if it was present in a country's economic system or a muddy puddle. We could build robots with real artificial consciousness, and there would be grounds for hope for people who want to scan their brains into a computer and extend their consciousness after their biological death (Chalmers 2010; Kurzweil 1999).

The experimental results that were discussed in the previous section are consistent with H1, but they do not prove it because they are also consistent with a pattern of neural activity being a correlate of consciousness. This paper will investigate whether experiments could in principle be carried out that could unambiguously demonstrate that H1 is true.

3. What is Information?

To test whether information patterns are correlated with consciousness we need to devise experiments that can separate informational correlates of consciousness from their implementation in a particular substrate. The first stage in this work is the development of a better understanding of the nature of information that will enable us to unambiguously identify information states in the brain. One way of approaching this is through Floridi's (2009) distinction between dedomena, data and information. According to Floridi's formulation of the general definition of information, σ is an instance of information, understood as semantic content, if and only if:

1) σ consists of n data, for $n >= 1$.

2) The data are well-formed.

3) The well formed data are meaningful.

This definition of *information* is based on the notion of *data*, which depends on pre-theoretical differences in the physical world known as *dedomena*. Each of these terms will now be explained in more detail.

3.1 Dedomena

Dedomena are pre-theoretical differences in the physical world that exist prior to human measurement and cannot be experienced independently of an interpretation that is applied to the world. They are the world as it is itself: a condition of possibility of measureable data - something like Kant's *noumena* or Locke's *substance*. For instance, dedomena make the measureable difference between higher and lower charge in a battery possible, and this type of measureable difference between physical states can be used to create higher levels of data, such as symbols.

3.2 Data

Data are non-uniformities in the world that are measured at a *level of abstraction* (LoA), which defines the scope and type of data in a system. As Floridi puts it: "... data are never accessed and elaborated (by an information agent) independently of a *level of abstraction* (LoA) ... A LoA is a specific set of typed variables, intuitively representable as an interface, which establishes the scope and type of data that will be available as a resource for the generation of information." (Floridi 2009, p. 37). Section 4 describes Floridi's theory of levels of abstraction in more detail and explains how data can be measured at different levels of abstraction in the brain.

3.3 Information

Information is defined by Floridi (2009) as well-formed meaningful data. Well-formed data is structured according to a set of rules that specify the valid combinations, how data can be written and read, and so on. For example, well-formed English text conforms to the rules of syntax and grammar and it is organized according to the conventions of reading from left to right and from top to bottom. Objects in a well-formed illustrative diagram typically conform to the rules of perspective and lighting.

The question about what makes data meaningful is more difficult. One of Floridi's suggestions is that meaningful data is a combination of data and queries. For example, the proposition "The earth only has one moon" can be interpreted as a piece of meaningful data in which the semantic content is the question "Does the Earth only have one moon?" and the answer "yes" is a single bit of data. Starting with the work of Shannon (1948), there has been an extensive amount of work on the communication of information, describing the entropy of an information source, the mutual information between two devices and the maximum rate of communication over a channel. However, as Floridi points out, Shannon's mathematical theory of communication (MTC) is a theory about data transmission, not information transmission, because it does not take the meaning of the messages into account: "... since MTC is a theory of information without meaning (not in the sense of meaningless, but in the sense of not yet meaningful), and since we have seen that [information – meaning = data], 'mathematical theory of data communication' is a far more appropriate description of this branch of probability theory than 'information theory'." (Floridi 2009, p. 33). This suggests that Tononi's theory of "information integration", which relies on Shannon's MTC, is more accurately described as a theory of data integration, unless it can be shown that the integrated data carries semantic content (Gamez 2011).

Before we can carry out experiments on the informational correlates of consciousness it is necessary to understand how we can measure meaningful data in the brain. One way of doing this would be to identify brain states that are functionally connected to states of the environment.[7] These states would be interpreted as meaningful because they are tracking features of the environment that are relevant to the organism. For example, some neurons are sensitive to bars of light (Hubel and Wiesel 1959) and others selectively respond to famous people (Quiroga et al. 2005). The problem with this approach is that it is impossible to probe the brain for all of its representational states. Even if we could expose the system to every known object, the brain might be hardwired to respond to things that we cannot even imagine. For example, suppose that a neuron in a Neanderthal's brain selectively responded to televisions - a Neanderthal neuroscientist who was not looking for a 'television neuron' would regard it as redundant or mis-wired. I have suggested elsewhere that we might be able to tackle this problem by analyzing the way in which the senses structure information (Gamez 2012b). How this would work for complex objects is an open question.

3.4 Dedomena, Data or Informational Correlates of Consciousness?

Which is the best candidate for a correlates-based approach to consciousness? While dedomena are the most objective aspect of the brain, since they cannot be directly measured it is impossible to prove that they are correlated with consciousness. The difficulties with identifying meaningful data (see previous section) will make it hard to prove that there are informational correlates of consciousness. This leaves the possibility that there could be data correlates of consciousness. This can be expressed as a revised version of H1:

> **H2**. One or more CC sets consist solely of data patterns. The substrate in which the data patterns are held is not part of these CC sets. A data pattern that was shown to be correlated with consciousness in one substrate (for example, the brain), would continue to be correlated with consciousness if it was implemented in a different substrate, such as a silicon chip.

Since Tononi's information integration theory of consciousness is actually a data integration theory, H2 fits in better with previous work on the correlates of consciousness than H1. The rest of this article will focus on the experimental testing of H2.

4. Measurement of Data

This section looks at how data can be measured at different levels of abstraction in the brain, the spatiotemporal patterns that can be identified in this data, and how these can be compared at different levels of abstraction.

4.1 Levels of Abstraction

Floridi (2008) theory of levels of abstraction is based on his definitions of typed variables and observables, which are organized into levels of abstraction in the following way:

- *Typed Variable*. A variable that holds one type of data. For example, variables can represent rational numbers or sets of colours. Two typed variables are regarded as equal if their variables have the same name and their types are equal sets.

[7] Functional connectivity (a deviation from statistical independence between A and B) is typically contrasted with anatomical connectivity (a physical link between A and B) and from effective connectivity (a causal link from A to B).

- *Observable*. A typed variable and a feature of a system that the observable corresponds to. The definition of an observable reflects a particular view or attitude towards the system that is being studied. Simple examples are the length of a snake and the voltage of a battery. Two observables are regarded as equal if they model the same feature of the system and their typed variables are equal.
- *Level of abstraction*. A set of observables. For example, the heights of a group of people.

Floridi (2008) suggests how the *behaviour* of observables can be specified at a level of abstraction and how different levels of abstraction can be combined into *gradients of abstraction*. While Floridi stresses that the concept of a level of abstraction is purely epistemological, the selection of a level of abstraction is likely to be influenced by the observer's theories about the ontology of the system.

4.2 Data Measurements at Different Levels of Abstraction in the Brain

The following list gives some examples of sets of observables that form levels of abstraction in the brain:

1. Electromagnetic activity of dendrites, neurons and neuron groups.
2. Ratio between oxygenated and de-oxygenated blood in brain areas of different sizes.
3. Last time each neuron spiked in Unix time.
4. Number of spikes of each neuron in the last 5 seconds, the last 10 seconds.
5. Number of spikes of each neuron group in the last 5 seconds, the last 10 seconds.
6. Differences in spike timings between neurons in a population.
7. Whether the number of spikes in a neuron group in the last second is odd or even.
8. The colour of each neuron, blood vessel, glia, etc.
9. Distance of each neuron from the North Pole.
10. Distance of each neuron group from the North Pole.
11. Temperature of each neuron, blood vessel, glia, etc.
12. Number of neutrons in brain areas of different sizes.

These observables can be measured using a variety of methods, including EEG, fMRI, implanted electrodes and optogenetics. Some of the limitations of these techniques are discussed in Section 5.8.

Some of these levels of abstraction are sensitive to the way in which the brain is partitioned - for example, the spiking activity of each neuron group depends on how the brain is divided up into neuron groups. The brain can also be divided up using dynamic partitioning schemes that change over time. Some levels of abstraction can be hierarchically organized into nested levels – for example, neurons form populations of neurons that can be organized into neuron groups, and so on. Other levels of abstraction do not fit together in a neat nested way - for example, there will be no overlap between a data set based on spike timings and a data set based on the distance of each neuron from the North Pole, and it will be impossible to derive one data set from the other.[8]

The numbers that are extracted at each level of abstraction are tied to a particular measurement system - for instance, length data can be expressed in metric or Imperial units. There is nothing mandatory about the units that are used, and so levels of abstraction can be based on novel and arbitrary systems of measurement – for example, I could introduce a new logarithmic unit of length based on my own foot. Two levels of abstraction that are based on different units can extract completely different sets of numbers from a system, even if they are using the same variable type and measuring the same aspect of the same physical system. These data sets can have complex mathematical relationships between them.

Some people might argue that we should stick to standard scientific ways of extracting data from the brain, and ignore the more unusual options that have been mentioned in this section. However, if H2 is correct, then *any* data set that can be extracted from the brain could contain spatiotemporal patterns that are correlated with consciousness. This includes the more exotic examples that I have mentioned, which are all legitimate data sets that can be extracted from the brain.

[8] This lack of relationship between some levels of abstraction in the brain is likely to prevent Floridi's (2008) concept of a gradient of abstraction from being applied to the problems discussed in this paper.

4.3 Spatiotemporal Data Patterns

Data sets cannot form CC sets by themselves because they can be extracted from both the conscious and unconscious brain - it is the spatiotemporal patterns within the data sets that are potential correlates of consciousness. These patterns can appear when the brain is conscious and be absent from the unconscious brain. The types of data pattern that could be correlated with consciousness include simple spatial or spatiotemporal patterns (Figure 1a, Figure 1d) and patterns that are identified by applying mathematics or algorithms to the data (Figure 1b and Figure 1c).[9]

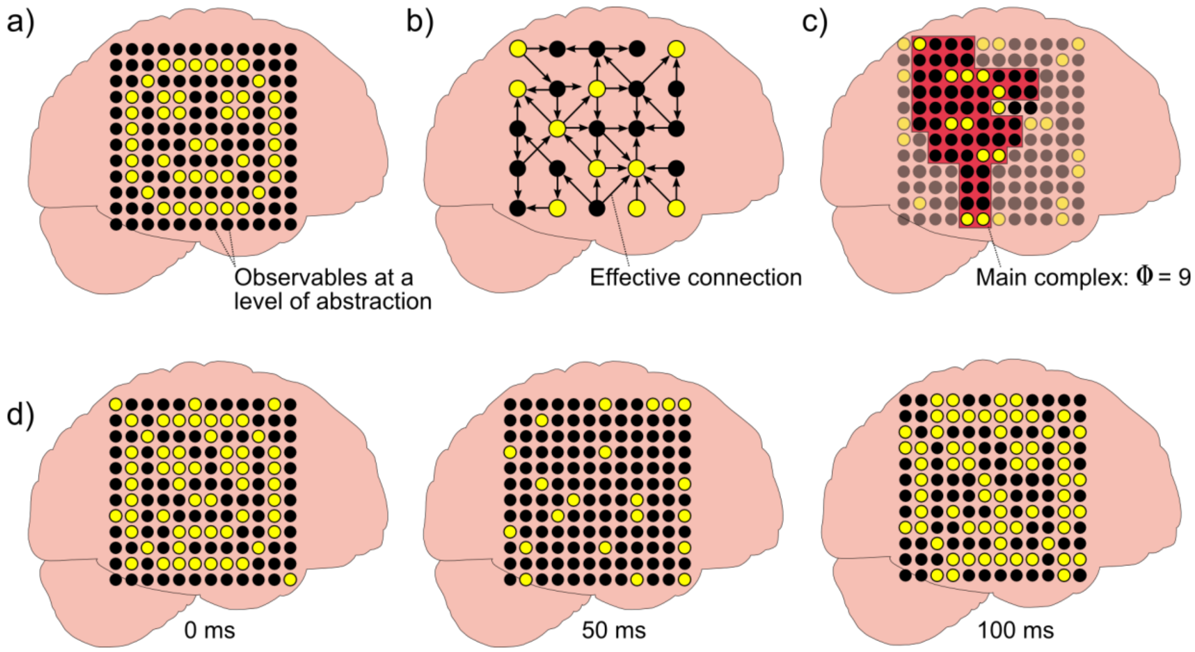

Figure 1. Examples of spatiotemporal patterns that could appear in data measured at a level of abstraction in the brain. The black circles are observables with binary type: black indicates that they have value 0; yellow indicates that they have value 1. These could, for example, be neuron firing events at a particular point in time. a) Simple spatial pattern. b) Connection pattern. Functional connections can be identified using mutual information; effective or causal connections can be inferred using Granger causality (Granger 1969) or transfer entropy (Schreiber 2000). Topological algorithms can be applied to the connection patterns to measure higher-order properties, such as knotty-centrality (Shanahan and Wildie 2012). c) Information/data integration pattern. An 'information integration' algorithm (see Appendix) is applied to the observables and outputs a part of the system called the main complex (highlighted in red), which is predicted to be the conscious part of the system, and a number, Φ, that is intended to correspond to the level or amount of consciousness. A high dimensional structure can also be generated that is supposed to correspond to the contents of consciousness (Balduzzi and Tononi 2009), which is not illustrated here. d) Synchronization pattern. The states of the observables are shown at intervals of 50ms. The observables in the face pattern are synchronized at 10 Hz.[10]

4.4 Comparison of Data Patterns at Different Levels of Abstraction

Experiments that attempt to test H2 need to compare data patterns at different levels of abstraction (see Section 5). This requires a definition of what it means for two data patterns to be the same or different, which will vary with the type of pattern. With spatial patterns, such as the face illustrated in Figure 1a, a simple intuitive approach would be to claim that two data patterns match if one pattern can be rescaled to map onto the other.[11] Two 'information integration' patterns (Figure 1c) will be the same if their main complexes are located in the same area of the brain, and if they have the same value of Φ and the same structure of 'information' relationships. With each type of

[9] Since data patterns are themselves data sets, data sets and data patterns could be combined into a hierarchy of levels of abstraction – what Floridi (2008) calls a gradient of abstraction. However, in experiments on the correlates of consciousness there is an important practical distinction between measurements of the system (using EEG, electrodes, fMRI, etc.) and the ways in which these measurements are processed by algorithms to identify patterns that might be correlated with consciousness. To preserve this distinction and minimise confusion, I will use 'data sets' to describe the different ways in which scientists extract measurements of the physical world at a level of abstraction, and talk about the possibility that patterns in these data sets might be correlated with consciousness (H2).

[10] Colour figures are available in the online version of this article.

[11] The mathematics of topology could be used to provide a more precise definition of spatial pattern matching, but this is beyond the scope of this paper.

pattern there is a possibility of partial or approximate match – for example, two 'information integration' patterns could be said to partially match if they had the same value of Φ but their main complexes were in different locations.

5. Experiments on the Data Correlates of Consciousness

This section examines whether experiments could in principle be carried out that could unambiguously demonstrate that H2 is true. The data patterns that are identified by 'information integration' algorithms (see Figure 1c) have been the main focus of this type of scientific research on consciousness, and so I will use these as my examples in the rest of this paper. Similar experimental issues are likely to apply to the other types of spatiotemporal data pattern that were mentioned in Section 4.3.

5.1 Methodology

An experiment on the data correlates of consciousness consists of the following steps:

1. Measure consciousness. For example, using verbal reports, button or lever pressing, or other behaviour.
2. Define a level of abstraction. This will include a procedure for converting measurements of the physical system into numbers and a specification of the spatial and temporal resolution of the measurements.
3. Measure system over a period of time to obtain the values of the variables in the observables.
4. Identify spatiotemporal patterns by putting the numbers into a data algorithm, such as the one developed by Balduzzi and Tononi (2008).
5. Carry out steps 1-4 on a conscious and unconscious brain, or on conscious and unconscious parts of a conscious brain, to identify correlations between the spatiotemporal data patterns and consciousness.

Some parts of this procedure are illustrated in Figure 2.

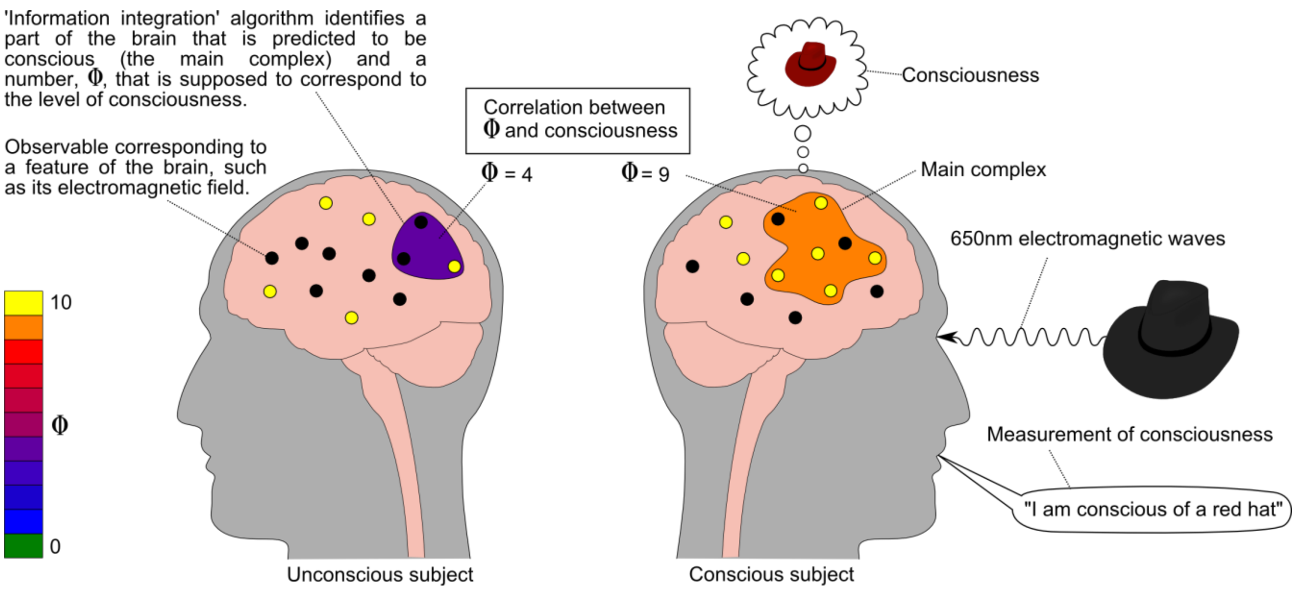

Figure 2. Experiment to identify spatiotemporal data patterns that are correlated with consciousness. This example is based on Balduzzi and Tononi's (2008) 'information integration' algorithm, which outputs an area of the system called the main complex that is supposed to be the conscious part of the system, and a number, Φ, that is intended to correspond to the level of consciousness. The value of Φ in the main complex is given by the colour scale, which will be used in subsequent illustrations. The subject on the left is unconscious and the 'information integration' algorithm outputs a low number - for example, a Φ value of 4. The subject on the right is conscious and the algorithm outputs a higher number, which is illustrated with a Φ value of 9. This shows that the output of the 'information integration' algorithm is correlated with the level of consciousness. A summary of Balduzzi and Tononi's (2008) algorithm is given in the appendix.

These experiments can only show that a data pattern is correlated with consciousness if they can meet the following requirements:

R1. *A correlation between a data pattern and consciousness does not depend on arbitrary choices made by the observer*. The presence or absence of consciousness is observer-independent, and so the features of the physical world that are correlated with consciousness must be observer-independent as well. Since the selection of a level of abstraction is necessarily a subjective choice, some method must be found for working across multiple levels of abstraction to eliminate or minimize this issue. H2 cannot be proved by focusing on neural activity and ignoring all other possible levels of abstraction.

R2. *A data correlate of consciousness is absent from the unconscious brain at all levels of abstraction*. It is not enough to select a level of abstraction and show that a data pattern at this level of abstraction is present when the brain is conscious and not present when the brain is unconscious. Since we are looking for a substrate-independent data correlate of consciousness, we have to show that a candidate data pattern is not present in the unconscious brain at any level of abstraction.

R3. *The correlation between a data pattern and consciousness is independent of the substrate that happens to instantiate it at a particular point in time*. Our experiments will have to resolve a potential ambiguity between two claims: (1) A pattern of *data* is correlated with consciousness; (2) A pattern of an *aspect of the physical world* is correlated with consciousness. Since experiments on the correlates of consciousness have to be carried out on the human brain (see Section 2.1), we will have to show that data patterns in the brain are correlated with consciousness independently of the physical structures in which the data is found.

The following sections will discuss some problems with experiments that attempt to prove H2 while meeting these requirements.

5.2 'Information Integration' in the Unconscious Brain

A first issue that needs to be addressed is that the current information/data algorithms are likely to return a positive result for the unconscious brain (see Figure 2), which would contradict the observation that we have *no* consciousness when we are unconscious (R2). This problem can be fixed by applying an empirically determined threshold to the output of the algorithm.[12]

5.3 Single Level Analysis

It might be thought that the subjectivity of levels of abstraction (R1) could be addressed by analyzing the brain at a single level, such as the neural level. We would identify correlations between spatiotemporal patterns at this level and consciousness and claim that we have identified data correlates of consciousness. However, this would only show that a particular *neuron* pattern was correlated with consciousness. To demonstrate that a data pattern is correlated we would have to show that it is present in the conscious brain and *absent from all of the possible levels of abstraction* when the brain is unconscious (R2). We would also have to show that the data pattern is not tied exclusively to the neural level, and can be correlated with consciousness at several different levels of abstraction (R3).

5.4 Cross Level Analysis

The subjectivity of levels of abstraction could be addressed by measuring the spatiotemporal data patterns at all possible levels of abstraction. One could then look for a data pattern that is present at one or more levels when the brain is conscious and absent from all levels when the brain is unconscious. The correlation between the data pattern and consciousness would no longer depend on the choice of a level of abstraction because all possible levels would be under consideration.

Another way of handling this problem would be to develop an algorithm that works across multiple levels of abstraction. For example, Balduzzi and Tononi's (2008) algorithm could be used to calculate Φ at each possible level of abstraction, and the level of abstraction at which Φ was maximized would become the final output that was potentially correlated with consciousness.[13] This is illustrated in Figure 3.

[12] A different way of addressing this problem would be to assume that a sleeping or anaesthetized brain is actually conscious, but unable to report or remember its consciousness. However, this would make contrastive experiments on the correlates of consciousness impossible because there will not be states of the system in which consciousness is known to be absent. I have discussed these issues elsewhere in 'The Measurement of Consciousness', currently under review. A draft is available on request.

[13] Tononi (2010) suggests this approach.

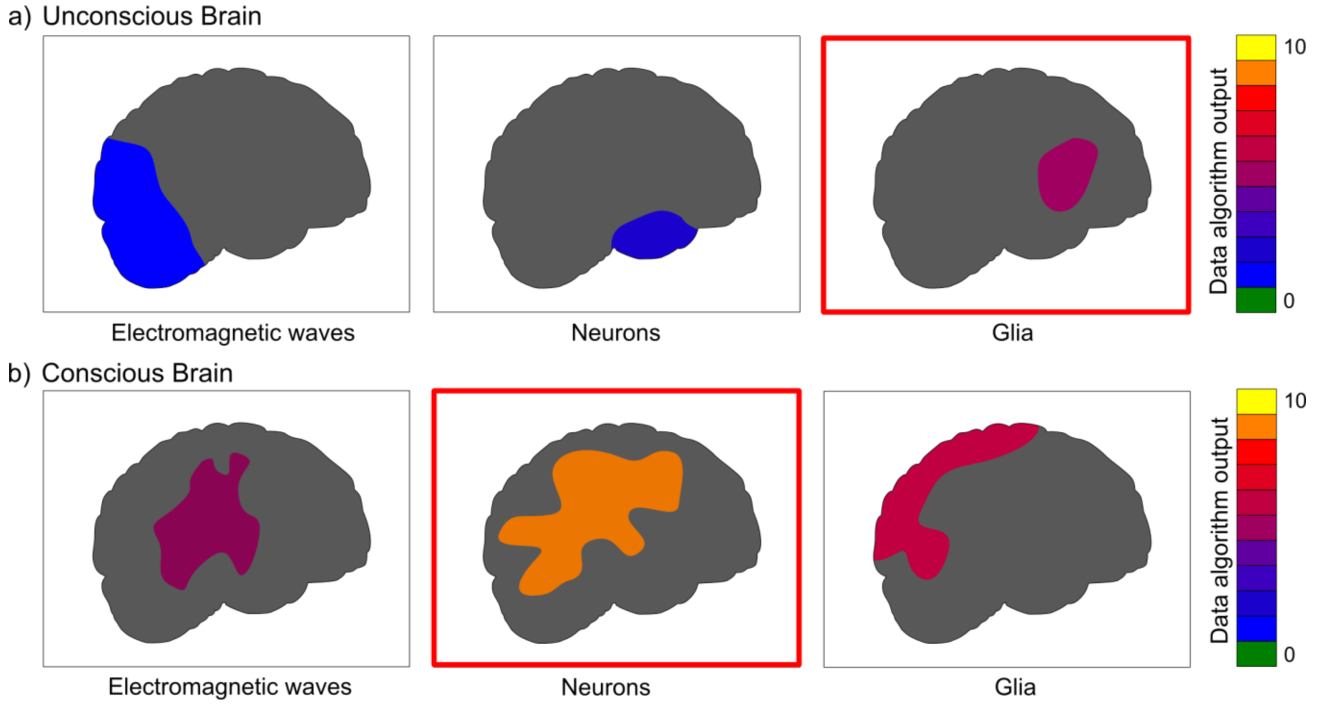

Figure 3. An experiment demonstrating a correlation between a spatiotemporal pattern that is output by a cross-level data algorithm and consciousness. In this and subsequent figures a red square is placed around the level of abstraction at which the algorithm reaches a maximum. The grey area of the brain does not have a value assigned to it because it is not predicted to be associated with consciousness. a) Cross level analysis of the unconscious brain in which the algorithm reaches a maximum value of 4 at the glia level. b) Cross-level analysis of the conscious brain in which the algorithm reaches a maximum value of 9 at the neuron level. This shows that there is a correlation between the output of the cross-level algorithm and consciousness. A threshold of 5 could be used to set the output of the algorithm to zero when the brain is unconscious (see Section 5.2).

One difficulty with cross-level analyses is that the lower levels of a system cannot be accurately measured, even in theory, because of Heisenberg's uncertainty principle. Other problems are discussed in the following sections.

5.5 Coincidence Between Levels of Abstraction

The potentially infinite number of levels of abstraction would cease to be an issue if the data patterns at different levels coincided – for example, if the data pattern at the level of electromagnetic waves was identical with the data patterns at the levels of neurons and glia. In this case it would not matter which level of abstraction was selected because they would all lead to the same result and the analysis of a potentially infinite number of levels of abstraction would be avoided (see Figure 4).

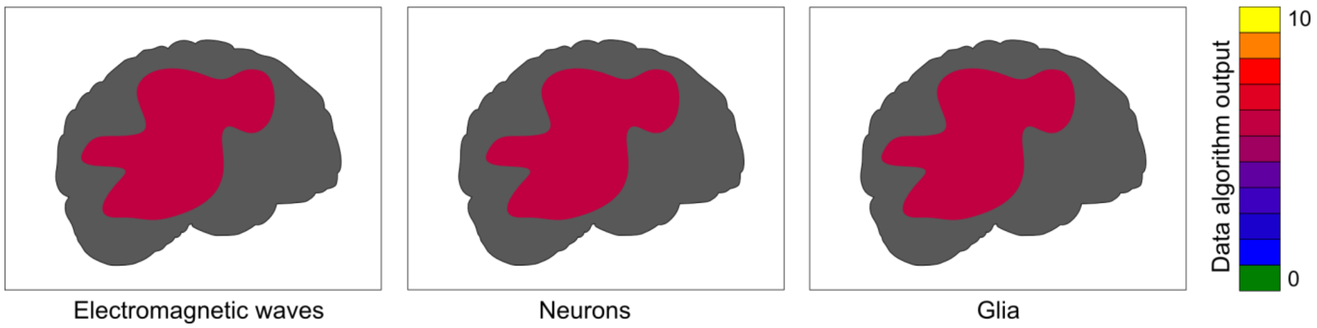

Figure 4. The same spatiotemporal pattern is output by the data algorithm at different levels of abstraction. A cross-level analysis is superfluous because the same result is obtained regardless of the level that is selected.

The problem with a coincidence between levels of abstraction is that it would prevent us from proving that consciousness was correlated with a substrate-independent data pattern (R3) - it would become an open question whether consciousness was correlated with a data pattern or with a pattern at a specific level. For example, if the neural activity pattern matched the electromagnetic wave pattern and the glia pattern, then it would be impossible to tell whether the data pattern at all of these levels correlated with consciousness, or whether the neural activity pattern was the true correlate and this just happened to coincide with everything else.

At present we have no grounds for believing that there will be a neat coincidence between the data patterns at different levels of abstraction. The examples of observables in Section 4 suggest that data patterns at different levels are likely to be substantially different.

5.6 Contradiction Between Levels of Abstraction

A cross-level analysis could lead to a situation in which the data algorithm produced the same numerical result for different areas of the brain at different levels of abstraction (see Figure 5). This would not affect the correlation between the output of a cross-level algorithm and consciousness, but it would mean that the choice of a level of abstraction affected the area of the brain that was predicted to be conscious. This will be referred to as a contradiction between levels of abstraction.

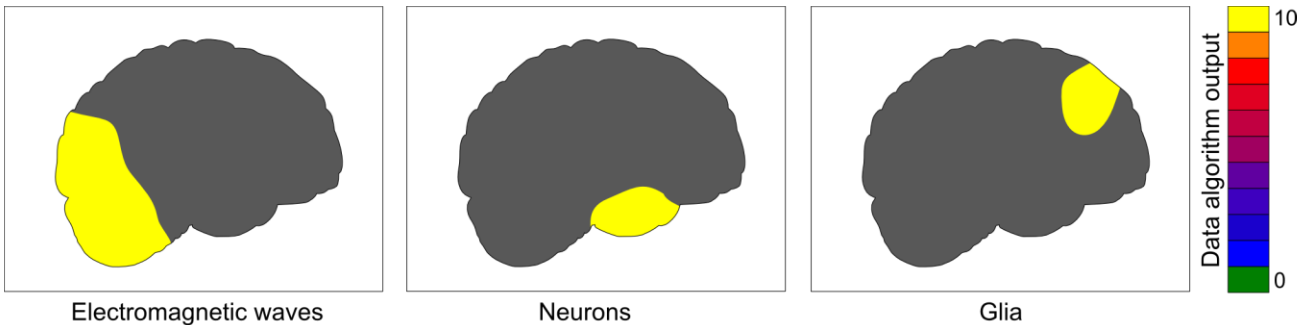

Figure 5. Contradiction between levels of abstraction. Cross-level maximisation does not work because the output of the data algorithm at different levels is the same. While this would not affect a correlation between the output of the algorithm and the level of consciousness, the area of the brain that is predicted to be correlated with consciousness varies with the level of abstraction.

If contradiction occurred, it is possible that only one level would match the conscious brain area, which would suggest that consciousness was correlated with a particular physical structure and not with a data pattern. It is also possible that individual areas could be removed without affecting consciousness - just as many columns can support a ceiling without any individual column being necessary. A third possibility is that the data patterns at each level of abstraction could be correlated with separate consciousnesses.

Since many of the brain's functions are carried out by particular brain areas, a contradiction between levels is likely to prevent us from matching the output of the algorithm to the contents of consciousness. To address this issue it might be possible to adjust the algorithm to avoid contradiction, or the union of the results could be taken across different levels. In practice this is unlikely to be a serious problem because the maxima at different levels will rarely be exactly equal.

5.7 Data Pattern or Physical Correlate?

It is possible that a cross-level algorithm's output will be consistently correlated with consciousness at one level of abstraction. For example, if Tononi's algorithm was applied to electromagnetic waves, neurons and glia, we might find that the neuron level had consistently higher Φ and was correlated with consciousness. While this result would be consistent with a correlation between data patterns and consciousness, it would also be consistent with a correlation between neural patterns and consciousness, and there would not be enough information to decide between these two possibilities.

It might be possible to show that the data pattern was independent of the substrate (R3) if one level correlated with consciousness at one time and another level correlated with consciousness at another time, either in the same or different subjects. It could then be claimed that there are data correlates of consciousness, and not just physical correlates at one particular level. This would only work if the levels did not coincide, if the levels did not contradict each other, and if the data pattern that was correlated with consciousness naturally shifted across levels (see Figure 6). Such a shift between levels would not entirely eliminate the physical interpretation because it could still be claimed that consciousness occurs whenever a particular neuron *or* glia pattern is present. The substrate-independent data interpretation would become more appealing as the number of levels increased.[14]

[14] It has been suggested that we could demonstrate substrate independence by replacing part or all of the brain with a functionally equivalent silicon chip (Chalmers 1995). This would hold the data patterns constant while the underlying substrate

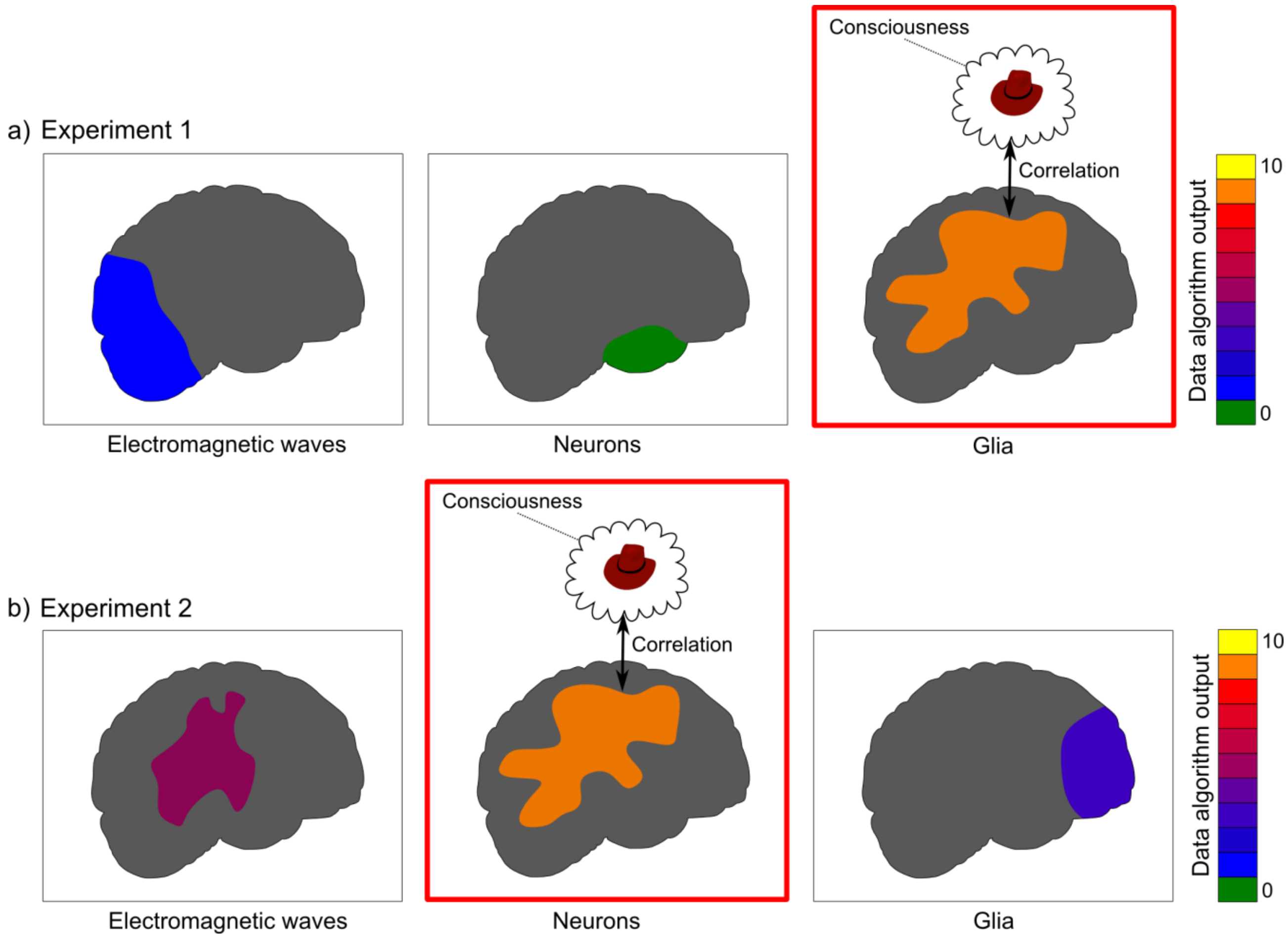

Figure 6. Demonstration that consciousness is correlated with a data pattern that can occur in different physical substrates. a) Experiment 1 (for example, the one shown in Figure 3) demonstrates that consciousness is correlated with the output of the data algorithm at the level of glia. a) A second experiment shows that the same consciousness is correlated with the same data pattern at the level of neurons. This suggests that consciousness is correlated with the data pattern and not with a pattern in a particular physical substrate.

5.8 Methods for Measuring the Brain

At present we have very limited access to the brain, which makes it very difficult to work across multiple levels of abstraction. Some of the current technologies are as follows:

- *fMRI*. A few thousand voxels; each corresponds the average activity of ~50,000 neurons over a few seconds.
- *EEG*. ~300 electrodes measure the brain's electrical field, typically through the scalp. Very poor spatial resolution; good temporal resolution.
- *Electrode arrays*. ~300 electrodes implanted in the brain with good spatial and temporal resolution. Only rarely used on human subjects.
- *Optogenetics*. It is now possible to record from 100,000 neurons in a zebrafish larvae in close to real time (Ahrens et al. 2013). It is more challenging to use optogenetics in mammalian brains and it cannot be used on human subjects.

---

was changed. If the system continued to be conscious, then we could claim that consciousness was correlated with data patterns, and not with a pattern in a particular physical substrate. The problem with this experiment is that the subject would behave in exactly the same way before and after the implantation of the chip. The only thing that could change during the experiment is our interpretation of their reports about consciousness. Either we would continue to believe that their reports were measurements of consciousness or we would believe that the substitution had removed the subject's capacity for consciousness and that the reports were zombie behaviour. Since the outcome of the experiment depends on the experimenter's prior assumptions about the brain, it would not be a valid demonstration that there are substrate-independent data correlates of consciousness. A more detailed discussion of this point is given by Gamez (2012a).

Out of these technologies, only optogenetics holds any promise of being able to work across multiple levels of abstraction in a mammalian brain. For example, we could use optogenetics to measure the activity of large numbers of individual neurons, and then calculate the neural synchronization patterns, the rate of neuron firing, the activity of neuron groups, and it might be possible to make reliable inferences about local field potential and EEG. However, many levels of abstraction in the brain (for instance, the sub-neural structures) are likely to remain inaccessible for the foreseeable future.

A second experimental issue is that some of the data algorithms depend on knowledge of the system's underlying causal structure, which is not directly available in data from the brain. Although causal relationships can be inferred from brain measurements using transfer entropy (Schreiber 2000) or Granger causality (Granger 1969), the inferred causal relationships might not match the actual causal relationships.

A third problem is that if we could extract high resolution data from the brain, we would be unable to analyze it with our current computer power. This is particularly apparent with Tononi's algorithms, which have factorial dependencies that make them impossible to compute on systems of more than ~20 elements (Gamez and Aleksander 2011). This issue might be solved by an increase in computer power, and it is possible to develop fast approximations to some algorithms.

6. Discussion and Conclusions

This article has outlined the scientific research on the neural correlates of consciousness and the proposals that have been made about informational or data correlates of consciousness in the brain. Using Floridi's distinction between information and data I have argued that the 'information integration' theory of consciousness is more correctly described as a data integration theory of consciousness, and that data correlates of consciousness are the best place to start with this work. The idea of data correlates of consciousness has intuitive appeal and it would answer questions about the possibility of consciousness in artificial systems and our ability to upload our brains into a computer and extend our consciousness after our biological death.

Virtually all of the previous experimental work on the correlates of consciousness has focused on the neural correlates of consciousness. To show that a data pattern is correlated with consciousness we have to demonstrate that it is present when the brain is conscious, absent from all levels of abstraction in the unconscious brain (R2), and that the measurement of this data pattern does not depend on a subjective choice of a level of abstraction (R1). It also has to be shown that the data pattern is correlated with consciousness independently of the physical substrate in which it happens to be instantiated (R3).

While it is likely to be reasonably straightforward to find a data pattern that is only present when consciousness is present, it is going to be very challenging to show that a data pattern is absent at all levels of abstraction in the unconscious brain. This requires an exhaustive search through a potentially infinite number of levels of abstraction or a cross-level algorithm that can work across all possible levels. Since we are likely to have very limited access to the brain for the foreseeable future, it is extremely improbable that we will be able to meet this requirement in the short or longer term.

If we could by some miracle prove that a candidate data pattern was absent from all levels of abstraction in the unconscious brain, then we would still have to prove that the data pattern was correlated with consciousness independently of the substrate in which it happened to be instantiated. I have suggested that this could be done if data patterns naturally shift between levels of abstraction in the conscious brain, but without systematic experiments it is difficult to know how likely this is to occur. Some insight into this issue could be gained by simulating the brain at multiple levels of abstraction (molecules, ion channels, electromagnetic waves, glia, etc.) and investigating whether data patterns shift across the simulated levels. [15]

---

[15] While we are unlikely to conclusively prove that a data pattern, d_1, is correlated with consciousness in a way that is consistent with R1-R3, we might still be able to show that d_1 is present at many levels of abstraction when the brain is conscious and not present at many levels of abstraction when the brain is unconscious. In this case it could be claimed that d_1 is a likely correlate of consciousness or the 'best game in town' for a correlate of consciousness.

If experiments have shown that d_1 has some substrate independence (R3), then it would be reasonable to claim that there is empirical support for a correlation between a data pattern and consciousness, and we could treat all systems in which d_1 can be found as potentially conscious. However, if there is no evidence for the substrate independence of d_1, then the experiments could also be interpreted to show that a pattern in a particular physical structure in the brain is correlated with consciousness - for example, a neural pattern or pattern in electromagnetic waves. In this case we should suspend judgement about whether d_1 forms a CC set by itself until evidence for substrate independence has been found.

These problems suggest that the prospects are bleak for scientifically proving that information or data patterns are correlated with consciousness in the brain. More progress is likely to be made if we focus on correlations between consciousness and spatiotemporal patterns at particular levels of abstraction. For example, while it is likely to be impossible to prove that ‘information’ integration is correlated with consciousness, the preliminary experimental work outlined in Section 2.2 suggests that ‘information integration’ algorithms might be a promising way of identifying *neural* patterns that are correlates of consciousness.

Acknowledgements

I would like to thank Barry Cooper and the John Templeton Foundation for supporting this work (Project ID 15619: 'Mind, Mechanism and Mathematics: Turing Centenary Research Project'). I would also like to thank Anil Seth and the Sackler Centre for Consciousness Science at the University of Sussex for hosting me as a Research Fellow during this project. I am grateful to the referees for their helpful comments, which have considerably improved this paper.

Appendix 1: Balduzzi and Tononi’s ‘Information Integration’ Algorithm

Balduzzi and Tononi’s (2008) ‘information integration’ algorithm is applied to a network of elements that are in a particular state. These elements are particular parts or aspects of a system that are linked to observables at a level of abstraction. The states of the observables are the values in the typed variables that change as the measured aspect of the system changes. Different levels of abstraction lead to different sets of observables that correspond to different elements in the system. Although this is actually a data integration algorithm (see Section 3.3), the summary in this appendix uses the same terminology as the original paper.

Balduzzi and Tononi’s (2008) algorithm uses relative entropy to measure the effective information that is generated by a subset of the elements when they enter a particular state. The relative entropy, $H[\,p\,||\,q]$, between probability distributions p and q is given by Equation 1:

$$H[p \,\|\, q] = \sum_i p_i \log_2 \frac{p_i}{q_i} \,. \tag{1}$$

When a set of elements are in a particular state, x_1, at time t_1 there is a certain probability that each possible state at t_0 led to the current state at t_1. The set of these probabilities is called the *a posteriori* repertoire or $p(X_0 \rightarrow x_1)$. The entering of the elements into state x_1 generates information because it defines the *a posteriori* repertoire of probabilities of the possible states that could have existed at t_0 and led to the current state. However, if the elements are in a state of maximum entropy and their states are entirely random, then each state x_1 could have been caused by any other state, and the fact that you are in x_1 tells you nothing about the previous state. In this case, the probability that each possible previous state of the elements caused the current state is the same, and this set of probabilities is known as the *a priori* repertoire, or $p^{max}(X_0)$. According to Balduzzi and Tononi (2008), the amount of information generated by a particular state can be measured using the relative entropy between the *a posteriori* repertoire associated with x_1 and the *a priori* repertoire, as expressed in Equation 2:

$$ei(X_0 \rightarrow x_1) = H[p(X_0 \rightarrow x_1) \,\|\, p^{\max}(X_0)] \,, \tag{2}$$

where $ei(X_0 \rightarrow x_1)$ is the effective information generated by the state x_1, $p(X_0 \rightarrow x_1)$ is the *a posteriori* repertoire and $p^{max}(X_0)$ is the *a priori* repertoire.

Equation 2 gives the effective information that is generated when some or all of the elements enter a particular state, but it does not tell us whether this information was generated by causal interactions among the elements, or whether it is the sum of the information generated by the elements acting independently. To answer this question Balduzzi and Tononi (2008) consider partitions of the system and calculate the relative entropy between the *a posteriori* repertoires of the parts considered independently and the *a posteriori* repertoire of the whole subset, as expressed in Equation 3:

$$ei(X_0 \rightarrow x_1 / P) = H\left[p(X_0 \rightarrow x_1) \middle\| \prod_{M^k \in P} p(M_0^k \rightarrow \mu_1^k) \right], \quad (3)$$

where $ei(X_0 \rightarrow x_1 / P)$ is the effective information of a particular partition, P, of the system into two or more parts, M^k is a part of the system, and μ^k is a state of M^k. To calculate $ei(X_0 \rightarrow x_1 / P)$, each part is considered as a system in its own right and the inputs from the other parts are treated as noise.

The *minimum information partition* is the subset or division of the subset across which the least information is integrated, and it is identified by comparing the normalized effective information values for each possible partition as well as the normalized effective information for the subset as a whole (known as the total partition, whose effective information is calculated using Equation 2). Normalization is needed during this comparison process because the effective information across a partition between a single element and a number of elements is typically less than an equal bipartition, and the effective information across many partitions is typically higher than the effective information across few partitions.[16] The un-normalized value of effective information for the minimum information partition is the Φ value of the subset, which is calculated for every possible subset of the system.

Balduzzi and Tononi (2008) define a *complex* as a subset that is not included in another subset with higher Φ. According to Balduzzi and Tononi, complexes are regions of the system where elements integrate the most information, and the Φ value of each complex corresponds to the amount of information that is integrated. The *main complexes* of the system are the complexes whose subsets have strictly lower Φ, and Tononi (2004, 2008) claims that main complexes are the conscious parts of the system. A correlation-based interpretation of information integration would interpret the main complexes as the parts of the system that are predicted to be correlated with consciousness.

References

Ahrens, M B, Orger, M B, Robson, D N, et al (2013) Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nat Methods 10(5): 413-420

Aru, J, Bachmann, T, Singer, W, et al (2012) Distilling the neural correlates of consciousness. Neuroscience & Biobehavioral Reviews 36(2): 737-746

Balduzzi, D, Tononi, G (2008) Integrated information in discrete dynamical systems: motivation and theoretical framework. PLoS Comput Biol 4(6): e1000091

Balduzzi, D, Tononi, G (2009) Qualia: the geometry of integrated information. PLoS Comput Biol 5(8): e1000462

Barrett, A B, Seth, A K (2011) Practical measures of integrated information for time-series data. PLoS Comput Biol 7(1): e1001052

Block, N (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. Behavioral and Brain Sciences 30(5-6): 481-499

Casali, A G, Gosseries, O, Rosanova, M, et al (2013) A theoretically based index of consciousness independent of sensory processing and behavior. Sci Transl Med 5(198): 198ra105

Chalmers, D (1995) Absent Qualia, Fading Qualia, Dancing Qualia. In: Metzinger, T (ed.) Conscious Experience. Imprint Academic, Thorverton

Chalmers, D (2000) What Is a Neural Correlate of Consciousness? In: Metzinger, T (ed.) Neural Correlates of Consciousness. MIT Press, Cambridge, Massachusetts

Chalmers, D (2010) The Singularity: A Philosophical Analysis. Journal of Consciousness Studies 17: 7-65

Chalmers, D (2011) A Computational Foundation for the Study of Cognition. Journal of Cognitive Science 12(4): 323-357

de Graaf, T A, Hsieh, P J, Sack, A T (2012) The 'correlates' in neural correlates of consciousness. Neuroscience & Biobehavioral Reviews 36(1): 191-197

Dehaene, S, Naccache, L, Cohen, L, et al (2001) Cerebral mechanisms of word masking and unconscious repetition priming. Nature Neuroscience 4(7): 752-758

Ferrarelli, F, Massimini, M, Sarasso, S, et al (2010) Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. Proc Natl Acad Sci U S A 107(6): 2681-2686

Floridi, L (2008) The Method of Levels of Abstraction. Minds and Machines 18(3): 303-329

[16] Normalization can introduce instabilities in the Φ value of the subset – see Barrett and Seth (2011).

Floridi, L (2009) Philosophical Conceptions of Information. Lecture Notes in Computer Science 5363: 13-53

Gamez, D (2011) Information and Consciousness. Etica & Politica / Ethics & Politics, XIII(2): 215-234

Gamez, D (2012a) Empirically Grounded Claims about Consciousness in Computers. International Journal of Machine Consciousness 4(2): 421-438

Gamez, D (2012b) From Baconian to Popperian Neuroscience. Neural Systems and Circuits 2(1): 2

Gamez, D, Aleksander, I (2011) Accuracy and performance of the state-based Φ and liveliness measures of information integration. Consciousness and Cognition 20(4): 1403-1424

Granger, C (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37: 424-438

Hadley, M W, McGranaghan, M F, Willey, A, et al (2012) A new measure based on degree distribution that links information theory and network graph analysis. Neural Systems and Circuits 2(1): 7

Hubel, D H, Wiesel, T N (1959) Receptive fields of single neurones in the cat's striate cortex. J Physiol 148: 574-591

Koch, C (2004). The quest for consciousness: a neurobiological approach. Roberts & Company, Englewood

Kurzweil, R (1999). The age of spiritual machines. Viking, New York

Lee, U, Mashour, G A, Kim, S, et al (2009) Propofol induction reduces the capacity for neural information integration: implications for the mechanism of consciousness and general anesthesia. Consciousness and Cognition 18(1): 56-64

Logothetis, N K (1998) Single units and conscious vision. Philos Trans R Soc Lond B Biol Sci 353(1377): 1801-1818

Massimini, M, Boly, M, Casali, A, et al (2009) A perturbational approach for evaluating the brain's capacity for consciousness. Prog Brain Res 177: 201-214

Molyneux, B (2010) Why the Neural Correlates of Consciousness Cannot be Found. Journal of Consciousness Studies 17(9-10): 168-188

Noë, A, Thompson, E (2004) Are There Neural Correlates of Consciousness? Journal of Consciousness Studies 11(1): 3-28

Popper, K R (2002). The Logic of Scientific Discovery. Routledge, London

Quiroga, R Q, Reddy, L, Kreiman, G, et al (2005) Invariant visual representation by single neurons in the human brain. Nature 435(7045): 1102-1107

Schreiber, T (2000) Measuring information transfer. Physical Review Letters 85(2): 461-464

Seth, A K, Izhikevich, E, Reeke, G N, et al (2006) Theories and measures of consciousness: an extended framework. Proceedings of the National Academy of Sciences of the United States of America 103(28): 10799-10804

Shanahan, M, Wildie, M (2012) Knotty-centrality: finding the connective core of a complex network. PLoS One 7(5): e36579

Shannon, C E (1948) A Mathematical Theory of Communication. The Bell System Technical Journal 27: 379–423, 623–656

Teasdale, G, Jennett, B (1974) Assessment of coma and impaired consciousness. A practical scale. Lancet 2(7872): 81-84

Tononi, G (2004) An information integration theory of consciousness. BMC Neurosci 5: 42

Tononi, G (2008) Consciousness as integrated information: a provisional manifesto. Biological Bulletin 215(3): 216-242

Tononi, G (2010) Information integration: its relevance to brain function and consciousness. Arch Ital Biol 148(3): 299-322

Tononi, G, Koch, C (2008) The neural correlates of consciousness: an update. Annals of the New York Academy of Sciences 1124: 239-261

Tononi, G, Sporns, O (2003) Measuring information integration. BMC Neurosci 4: 31

Tononi, G, Sporns, O, Edelman, G M (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system. Proc Natl Acad Sci U S A 91(11): 5033-5037

Wilkes, K V (1988) ___, yìshì, duh, um, and consciousness In: Marcel, A J, & E Bisiach (eds.), Consciousness in Contemporary Science. Clarendon Press, Oxford